Journal of Industrial Microbiology and Biotechnology, 2022, 49, kuab080

https://doi.org/10.1093/jimb/kuab080
Advance access publication date: 3 November 2021
Biocatalysis – Review

Bacterial xylan utilization regulons: systems for coupling depolymerization of methylglucuronoxylans with assimilation and metabolism

Virgina Chow[1], Guang Nong[1], Franz J. St John[2], Neha Sawhney[1], John D. Rice[1], James F. Preston[1]

[1]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA
[2]Forest Products Laboratory, Institute for Microbial and Biochemical Technology, USDA Forest Service, Madison, WI 53726, USA
Correspondence should be addressed to: Franz J. St John. Phone: +608 231 9438. E-mail: fjstjohn@gmail.com; franz.j.st.john@usda.gov

Abstract: Bioconversion of lignocellulosic resources offers an economically promising path to renewable energy. Technological challenges to achieving bioconversion include the development of cost-effective processes that render the cellulose and hemicellulose components of these resources to fermentable hexoses and pentoses. Natural bioprocessing of the hemicellulose fraction of lignocellulosic biomass requires depolymerization of methylglucuronoxylans. This requires secretion of endoxylanases that release xylooligosaccharides and aldouronates. Physiological, biochemical, and genetic studies with selected bacteria support a process in which a cell-anchored multimodular GH10 endoxylanase catalyzes release of the hydrolysis products, aldotetrauronate, xylotriose, and xylobiose, which are directly assimilated and metabolized. Gene clusters encoding intracellular enzymes, including α-glucuronidase, endoxylanase, β-xylosidase, ABC transporter proteins, and transcriptional regulators, are coordinately responsive to substrate induction or repression. The rapid rates of glucuronoxylan utilization and microbial growth, along with the absence of detectable products of depolymerization in the medium, indicate that assimilation and depolymerization are coupled processes. Genomic comparisons provide evidence that such systems occur in xylanolytic species in several genera, including *Clostridium*, *Geobacillus*, *Paenibacillus*, and *Thermotoga*. These systems offer promise, either in their native configurations or through gene transfer to other organisms, to develop biocatalysts for efficient production of fuels and chemicals from the hemicellulose fractions of lignocellulosic resources.

Keywords: Endoxylanase, Methylglucuronoxylans, Assimilation, *Paenibacillus*

Introduction

Interest in developing alternatives to energy derived from fossil fuels has led to the development of microbial biocatalysts for the production of fuel ethanol and chemical feedstocks from resources renewable through photosynthesis. Yeast fermentation of glucose derived from starch and sucrose has allowed industrial production of ethanol from corn, sorghum, and sugarcane, and a similar process may be applied to the glucose derived from the cellulose fraction of lignocellulosic biomass (Agbogbo & Coward-Kelly, 2008; Dien et al., 2003; Sedlak & Ho, 2004). With attention now directed toward the efficient bioconversion of the lignocellulosic fraction derived from wood and agricultural residues, as well as evolving energy crops, there is a need for efficient conversion of both the hemicellulose and the cellulose fraction of these resources (Chundawet et al., 2011; Himmel et al., 2007).

Development of Biocatalysts for Saccharide Fermentation

The ability of genetically malleable Gram-negative bacteria, for example *Escherichia coli* and *Klebsiella oxytoca*, to efficiently ferment pentoses and hexoses has led to the engineering of strains of these bacteria for bioconversion of pretreated lignocellulosics to biobased products, and made them attractive candidates for the industrial production of fuel ethanol and chemical feedstocks (Dien et al., 2003; Ingram et al., 1987, 1999; Jarboe et al., 2007; Shanmugam & Ingram, 2008). The Gram-negative *Zymomonas mobilis*, with its metabolic commitment to ethanol fermentation via the Entner–Douderoff pathway, has been engineered for pentose fermentation and may also be used for industrial production of ethanol from cellulosic resources as well as starch. Gram-positive biocatalysts, for example species of *Clostridium*, *Thermoanaerobacterium*, *Bacillus* and *Caldocellulodisruptor* spp., have also been developed for production of biofuels and chemicals (Lynd et al.,1999; Shanmugam and Ingram, 2008; Shaw et al., 2008; Straub et al., 2019; Zhang & Lynd, 2005). Yeast strains have been developed for xylose fermentation and may provide useful alternatives for the conversion of hemicellulose-derived pentoses to ethanol (Jeffries, 2006; Jeffries & Jin, 2004; van Maris et al., 2006). Some of the Gram-positive bacteria, for example *Clostridium*, *Thermoanaerobacterium*, and *Caldicellulodisruptor* spp., that have been evaluated as biocatalysts for production of ethanol secrete cellulases and/or xylanases that may contribute to polysaccharide depolymerization as well as fermentation (Lynd et al., 2002; Mai & Wiegel, 2000; Shaw et al., 2008; Straub et al., 2019).

Pretreatment of Lignocellulosics for Saccharification and Fermentation

For hardwood and crop residues, as well as candidate energy crops, for example poplar and switchgrass, the cellulose and hemicellulose fractions comprised of fermentable sugars represent 40–45% and 18–22% of the total biomass, respectively (Saha, 2003). Industrial protocols for fermentation

Received: July 13, 2021. **Accepted:** October 25, 2021.

of plant biomass have generally relied upon thermochemical pretreatments for solubilization and hydrolysis of the hemicellulose fraction, leaving the insoluble cellulose for enzyme-mediated conversion to fermentable glucose. Common practice has been treatment in dilute acid (0.5–1.5% sulfuric acid) at high temperature (120–140°C), followed by separation of released saccharides from the unhydrolyzed cellulose fraction before enzymatic digestion and fermentation. Alternatively, the cellulose fraction may be subjected to fungal cellulases in the presence of ethanologenic biocatalysts without removal of the acid hydrolysate, allowing the simultaneous saccharification and fermentation (SSF) of glucose derived from cellulose and pentoses from the hemicellulose fraction.

A structural polymer common to hemicellulose fractions, and predominant in those derived from crop residues, is *O*-acetyl-arabinoglucuronoxylan, a β-1,4-linked xylan in which D-xylopyranose residues are variably substituted at 2-*O* positions with α-4-*O*-methylglucuronic acid (MeGA), and at 2-*O* and/or 3-*O* positions with α-L-arabinofuranose or acetyl groups. In hardwoods, the hemicellulose lacks arabinose substitutions and is referred to as methylglucuronoxylan ($MeGAX_n$). The degree to which xylose residues contain MeGA substitutions may be as high as 1 in 6 in the xylans in wood, with lesser amounts in xylans from straw and grasses (Jacobs et al., 2001; Dien et al., 2006; Preston et al., 2003). During the acid pretreatment step, arabinofuranosyl and *O*-acetyl groups are removed, and the β-1,4-xylosidic linkages are cleaved, resulting in the release of free xylose that is then available for fermentation by ethanologenic strains of *E. coli* and *K. oxytoca*. The release of xylose residues that are substituted with α-1,2-methylglucuronate is incomplete, generating quantities of the aldobiuronate, 4-*O*-methyl-D-glucuronosyl-α-1,2 xylose (MeGAX), which cannot be fermented by the ethanologenic biocatalysts in current use. A number of thermochemical pretreatment protocols have been developed to maximize yields of fermentable pentoses, for example xylose, as well as minimize inhibitor formation, for example furfural, that are generated from the dehydration of xylose (Liu & Wyman, 2005; Lloyd & Wyman, 2005; Wyman et al., 2005). These protocols may release greater amounts of aldouronates and lesser amounts of fermentable xylose and xylobiose. The development of biocatalysts to efficiently depolymerize $MeGAX_n$ and directly assimilate and ferment all of the products of depolymerization would allow for a more cost-effective production of biofuels and chemicals from lignocellulosic resources.

Bacterial Candidates for the Bioconversion of Hemicellulose

Direct utilization of lignocellulosic polysaccharides by bacteria requires the secretion of glycohydrolases, including β-1,4-endoglucanases (cellulases) and β-1,4-endoxylanases. Additional enzymes, including arabinofuranosidases and esterases, may also be secreted to process xylans for depolymerization and assimilation (Kuhad et al., 1997). This capability is found in many species of Gram-positive bacteria. Several *Clostridium* spp. with their cell-associated cellulosomes allow secreted cellulases to release cellobiose at the cell surface that may be efficiently assimilated and fermented by these bacteria, as with *Clostridium thermocellum* (Zhang & Lynd, 2005). More recent studies with hyperthermophilic bacteria, for example *Dictyoglomus thermophilum* and *Caldicellulodisruptor* spp., have provided biocatalysts for the direct conversion of cellulosic biomass to targeted fermentation products (Bergquist et al., 2014; Straub et al., 2019).

Many species of bacteria and fungi secrete endoxylanases and other glycohydrolases that participate in the depolymerization of the polysaccharides that comprise the hemicellulose fraction of lignocellulosics. The extracellular depolymerization of $MeGAX_n$ is catalyzed by endoxylanases in the glycohydrolase families GH30 (previously designated GH5) (Biely et al., 2015; Hurlbert & Preston, 2001; St John et al., 2006b, 2010; Vršanská et al., 2007), GH10, and GH11 (Biely et al., 1997; Collins et al, 2005; Preston et al., 2003; St John et al., 2006a). Members of each of these families catalyze the cleavage of a xylosidic bond by acid–base catalysis in which a glutamate residue serves as a nucleophile and a second glutamate serves as a proton donor, resulting in a double displacement reaction that retains the anomeric β-configuration. Of these enzymes, only the multimodular cell-associated GH10 endoxylanases, which catalyze the release of xylobiose, xylotriose, and the aldouronate $MeGAX_3$ in which the MeGA is linked to the nonreducing terminus of xylotriose, generate products all of which may be directly assimilated by xylanolytic bacteria. A diagram depicting the structure of the *O*-acetyl-arabinomethylglucuronoxylan and the processing provided by the different options is presented in Fig. 1.

Paenibacillus sp. JDR-2 as a Model for the Coupling of the Depolymerization of $MeGAX_n$ With Direct Assimilation and Metabolism

An aggressively xylanolytic bacterium, *Paenibacillus* sp. JDR-2, isolated from decaying hardwood, secretes a multimodular GH10 xylanase (XynA1) anchored to the cell surface (St John et al., 2006a). Genes encoding transcriptional regulators, ABC transporter proteins, a GH67 α-glucuronidase (AguA), a GH10 xylanase catalytic domain (XynA2), and a putative GH43 β-xylosidase/α-arabinofuranosidase (XynB), have been located within a contiguous sequence in the genome of this bacterium. All of these genes, as well as that of the XynA1 xylanase, are coordinately responsive to xylan induction and glucose repression, and collectively comprise part or all of a xylan-utilization regulon (Chow et al., 2007; Sawhney et al., 2015, 2016). Physiological and enzymatic studies support a process in which the $MeGAX_3$ and xylooligosaccharides (Fig. 1) released by XynA1 anchored to the cell surface are directly assimilated and metabolized. Growth rates and yields with $MeGAX_n$ as substrate and the absence of detectable xylanase hydrolysis products in the medium suggest that depolymerization and assimilation are coupled (Nong et al., 2009). The genes that comprise the xylan-utilization regulon in the genome of *Paenibacillus* sp. JDR-2 were used to identify orthologous genes in the genomes of other bacteria. The arrangements as well as sequences of these genes were examined to identify themes to predict a molecular basis for the formation of enzymes that contribute to the coupling of depolymerization to assimilation and the metabolism of the assimilated products. The identification of regulatory genes has also provided insight into the genetic requirements for applying these regulons for the efficient conversion of $MeGAX_n$ to targeted products.

Genomic Organization and Gene Function of the Xylan-Utilization Regulon of *Paenibacillus* sp. JDR-2

The direct and complete utilization of methylglucuronxylan may be accomplished by bacteria that secrete a GH10 endoxylanase,

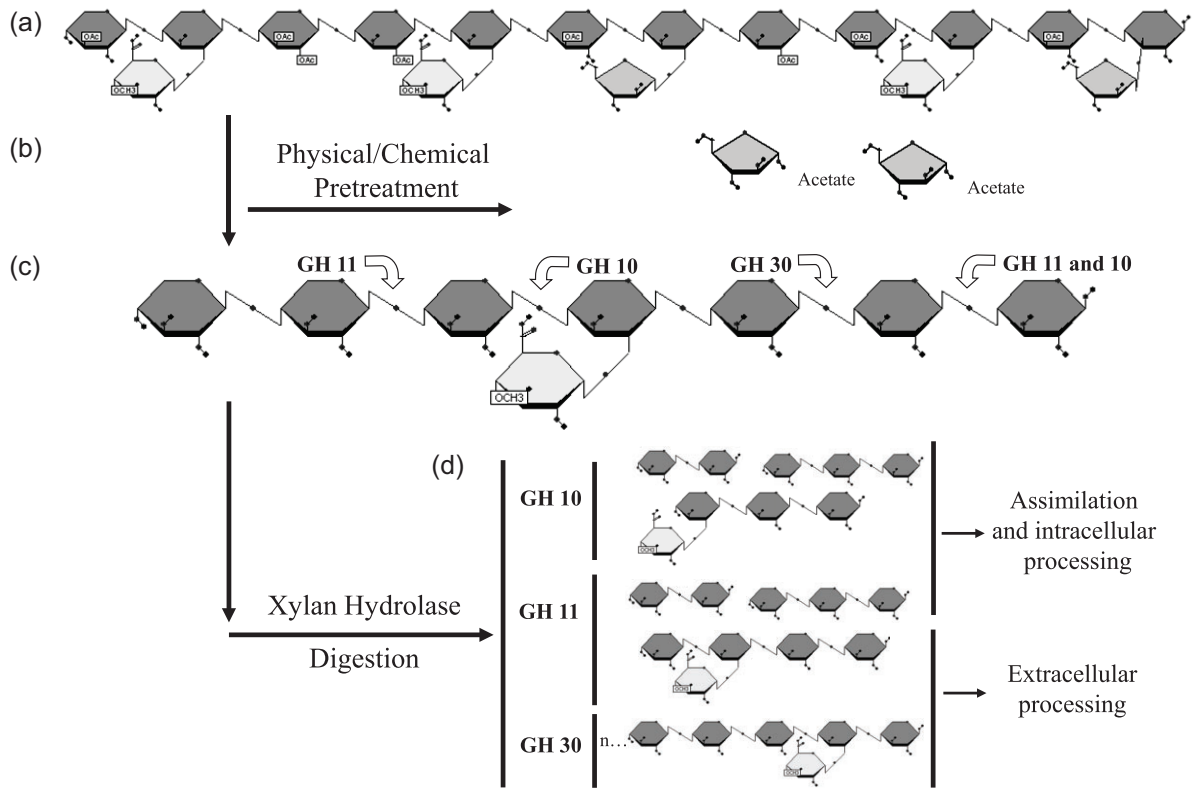

Fig. 1 Pretreatment of *O*-acetyl-arabinoglucuronoxylan and enzymatic depolymerization of 4-*O*-methylglucuronoxylan (MeGAX$_n$). (a) Structural presentation typical for hardwood xylan as a linear chain of β-1,4-linked xylopyranose residues variably substituted at the 2-*O* positions with α- 4-*O*-methyl-D-glucuronopyranosyl residues, and at the 2-*O* and 3-*O* positions with α-L-arabinofuranosyl residues or acetyl ester groups. (b) Thermochemical pretreatments to release arabinofuranose and acetate from soluble MeGAX$_n$. (c) Digestion of MeGAX$_n$ with endoxylanases of glycohydrolase family GH10, GH11, or GH30. (d) Release of aldouronates and xylooligosaccharides for direct assimilation or further extracellular processing.

and transporters that function to actively transport the aldouronate and xylooligosaccharide products of extracellular depolymerization (Biely et al., 1997; Preston et al., 2003). The GH10 endoxylanases may be identified by the characteristic amino acid sequence of a catalytic domain (295–412 amino acids in length for the examples in Table 1) that folds into a $(\beta/\alpha)_8$ barrel structure and directs the positioning of the catalytic nucleophile and proton donor glutamates for specific hydrolysis of the β-1,4 linkages of xylan. A structure for the recombinant XynA1 catalytic domain (CD) that derives from the cell-associated multimodular GH10 endoxylanases secreted by *Paenibacillus* sp. JDR-2 is presented in Fig. 2a (St John et al., 2012). This structure, which defines the binding site of MeGAX$_3$, is characteristic of the GH10 endoxylanase family and is typical of the XynA1 catalytic domains of the bacteria compared in Table 1.

A large number of the secreted GH10 endoxylanases have been shown to be multimodular, containing carbohydrate-binding modules (CBMs) that favor interaction with the cellulose and/or xylan regions of the lignocellulosic complex (Boranston et al., 2004; Ito et al., 2003; St John et al., 2006a). Many of these multimodular enzymes contain surface layer homology (SLH) domains that anchor the enzyme to the cell surface through molecular interactions with secondary cell wall polysaccharides of Gram-positive bacteria (Mesnage et al., 2000; Ozedemir et al., 2012; St John et al., 2006a; Zhao, Ali, et al., 2006; Zhao, Wamalwa, et al., 2006). The native XynA1 from *Paenibacillus* sp. JDR-2 is a multimodular GH10 xylanase that includes three N-terminal CBM22 modules, a single CBM9 module just toward the C-terminus to the GH10 catalytic domain, and C-terminal SLH domains. Characterization supports a function as a cell-anchored GH10 endoxylanase that catalyzes the depolymerization of MeGAX$_n$ with formation of MeGAX$_3$, xylotriose, and xylobiose, all of which are assimilated as they are formed (St John et al., 2006a, Nong et al., 2009). The multiple N-terminal CBM22 modules are predicted to selectively associate with glucan (i.e. xylan or β-glucan) (Xie et al., 2001) and the CBM9 module has been shown to be in tight coordination with reducing terminal sugars (Notenboom et al., 2001). From this assessment, the diagram presented in Fig. 2b reveals a natural directionality oriented toward the cell surface. This gene is represented by a single copy in the sequenced genome (Chow et al., 2012), as shown in Fig. 3.

The absence of aldouronates and xylooligosaccharides in the medium during growth of *Paenibacillus* sp. JDR-2 on MeGAX$_n$ predicts an active assimilation and metabolism process supported by genes encoding an intracellular GH67 AguA and xylosidases as well as transporter proteins. Previous studies on glucuronate utilization in *Geobacillus stearothermophilus* T6 have identified genes encoding these enzymes and implicated them in a pathway that includes the assimilation of aldotetrauronate MeGAX$_3$ and xylooligosaccharides by ABC transporter proteins followed by intracellular processing and metabolism (Shulami et al., 1999). Expression of genes derived from a cosmid library

Table 1. Comparison of genes comprising a xylan-utilization regulon in *Paenibacillus* sp. JDR-2 (PbsJDR-2) with orthologs from other xylanolytic bacteria. Bacteria were selected for comparison based upon the presence of genes encoding a GH67 α-glucuronidase (AguA), a multimodular GH10 xylanase (XynA1) with a secretory signal sequence, a unimodular intracellular GH10 xylanase catalytic domain (XynA2), a substrate-binding protein located within a contiguous sequence of genes encoding other components of an ABC transporter complex and a regulatory protein, all proximal to the *aguA* gene. Selections based upon these criteria are presented to include the percentage of their amino acid similarities (SI) to PbsJDR-2 as determined by blastp

	Genome	AguA			XynA1 CD[e]			XynA2			Sbp			AraC RR		
Organism[a]	GC%	SI[b]	AAs[c]	GC%[d]	SI	AAs	GC%	SI	AAs	GC%	SI	AAs	GC%	SI	AAs	GC%
PbsJDR-2	50.0	100	687	54.5	100	332	50.3	100	341	51.7	100	570	51.6	100	522	49.4
Clost H10	37.4	73	689	41.2	87	332	39.2	55	423	40.3	61	573	38.5	53	525	37.7
GeobMC10	51.0	78	693	53.2	47	369	55.1	78	346	51.4	72	561	52.3	64	533	55.6
GeobT6	47.1	75	679	47.2	51	412	45.1	77	331	46.6	46	441	44.8	37	259	44.0
Strep4680	70.7	66	680	73.0	49	295	68.3	55	358	69.4	42	549	66.5	47	270	80.0
TherMSB8	46.2	74	674	44.6	59	323	48.5	54	347	42.6	42	443	47.2	NA		

[a]The accessions of genome (DNA) and proteins listed in order: Genome, AguA, XynA1 CD, XynA2, Sbp (substrate-binding protein), and AraC RR (regulator of AraC family). *Paenibacillius* JDR-2 (PbsJDR-2): NZ_ABKS00000000, EDS50277, EDS53078, EDS50226, EDS50280, and EDS50282; *Clostridium cellulolyticum* H10 (Clost H10): NC_011898, YP_002504509, YP_002506636, YP_002505277, YP_002504518, YP_002504515; *Geobacillus* sp. Y412MC10 (GeobMC10): NZ_ABRG00000000, ZP_03038390, ZP_03036168, ZP_03040360, ZP_03037615, and ZP_03037929; *Geobacillus stearothermophilus* T6 (GeobT6): DQ868502, ABI49940, ABI49951, ABI49937, ABI49932, and ABI49931; *Streptomyces avermitilis* MA-4680 (Strep4680): NC_003155, NP_822625, NP_826161, NP_823272, NP_822629, and NP_825180; *Thermotoga maritima* MSB8 (TherMSB8): NC_000853, NP_227871, NP_227877, NP_227886, and NP_228 228. An ortholog-encoding AraC RR in *T. maritima* MSB8 was not found.
[b]The similarities of protein amino acid sequences (SI) were obtained by Blastp searching individual genomes for matches with translations of individual genes in *Paenibacillus* sp. JDR-2
[c]The number of amino acids (AAs) in the protein or protein domain.
[d]The percent composition of guanine and cytosine (GC%) contained in the DNA encoding the selected protein.
[e]XynA1 CD represents the catalytic domain defined within the complete sequence of the gene to exclude modules, linkers, and signal domains.

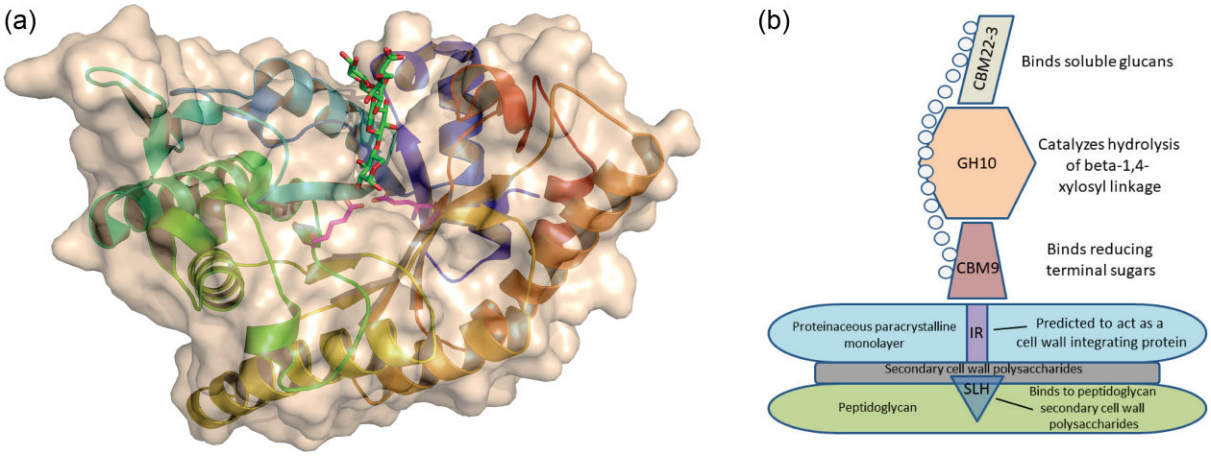

Figure 2. Presentation of the XynA1 xylanase from *Paenibacillus* sp. JDR-2. Structural model (a) of XynA1CD from *Paenibacillus* sp. JDR-2 (PDB accession: 3RDK) (St John et al., 2006a, 2012). The catalytic nucleophile (right, E262, corresponding to E780 in the native enzyme) and the catalytic acid/base catalyst (left, E138, corresponding to E656 in the native enzyme) are shown juxtaposed in the center straddling the position of the β-1,4-linkage to be hydrolyzed. Also, from this X-ray crystallographic structure model the glucuronoxylan hydrolysis limit product ligand aldotetrauronate is seen positioned into the glycone region of the XynA1CD substrate-binding cleft. The catalytic domain of XynA1 consists of amino acids 523–856 of the multimodular protein (UniProt/TrEMBL Q53I45). A cartoon showing the known modular order (b) for XynA1 from *Paenibacillus* sp. JDR-2. Given this order, and the module functional annotation this diagram proposes an inherent directionality to polymeric xylan hydrolysis and uptake.

of *Paenibacillus* sp. JDR-2 (Chow et al., 2007) as well as a full transcriptome (Sawhney et al., 2015, 2016) has defined a contiguous 15 kb sequence designated as an aldouronate-utilization gene cluster. Based upon homologies with genes defined in other bacteria, genes within this cluster encoded transcriptional regulators, ABC transporter proteins, a GH67 AguA, a GH10 xylanase catalytic domain (XynA2), and a putative GH43 β-xylosidase/α-arabinofuranosidase (XynB). All of these genes, in addition to *xynA1* encoding the multimodular xylanase, are coordinately responsive to xylan induction and glucose repression, and collectively comprise part or all of a xylan-utilization regulon (Chow et al., 2007, Sawhney et al., 2016). The catalytic properties of recombinant AguA and XynA2 have been defined as participating in the conversion of MeGAX$_3$ to xylose and xylobiose (Nong et al., 2009). Based upon the sequenced *Paenibacillus* sp. JDR-2 genome, there is a single *aguA* gene encoding a GH67 AguA and the end of the aldouronate gene cluster is separated by 1,300,949 bp from the start of the *xynA1* gene (Fig. 3). The coordinate regulation of the *xynA1* genes and those of the aldouronate-utilization cluster supports their roles within a xylan-utilization regulon.

Orthologous Genes and Genomic Organization in Other Bacteria

Genomic nucleotide sequences and protein/enzyme amino acid sequences were used to identify orthologs for all of the genes comprising the xylan-utilization regulon of *Paenibacillus* sp. JDR-2. For species within the bacterial phylum Firmicutes, the presence

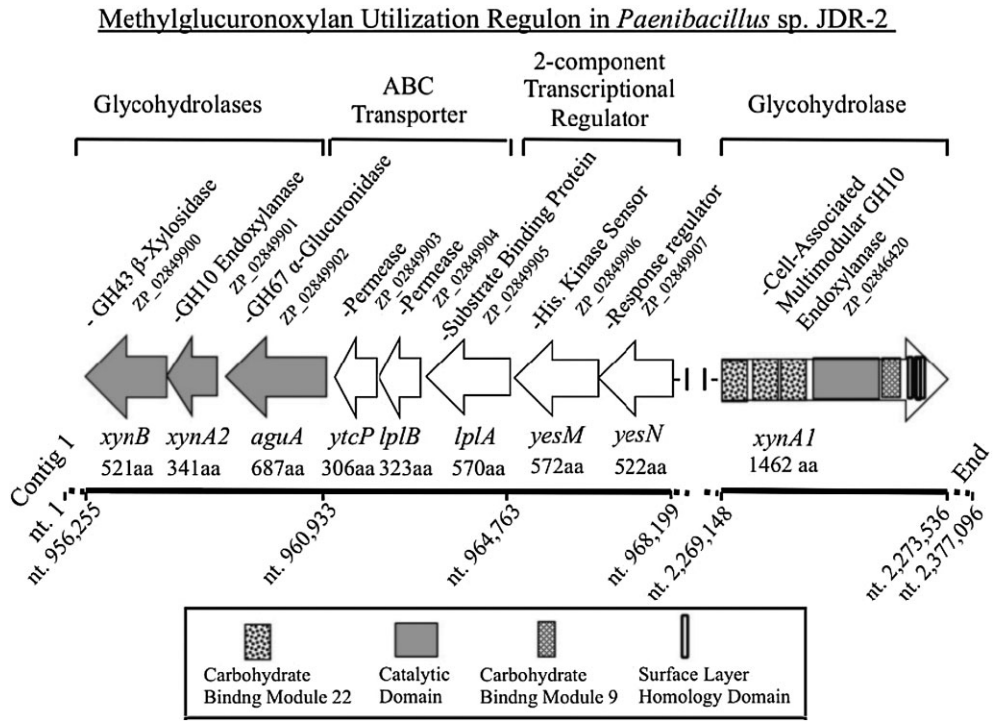

Figure 3. Genomic distribution of structural genes comprising the xylan-utilization regulon in *Paenibacillus* sp. JDR-2. Sequences were determined in genomic libraries (Chow et al., 2007; St John et al., 2006a) and confirmed in the draft sequence of the genome (DOE Joint Genome Institute, Genome Project 20399). The accession number to the nucleotide sequence of the operon containing *aguA* is ABKS01000022 and the accession number to the nucleotide sequence of *xynA1* is ABKS01000003, both available at the National Center for Biotechnology Information (NCBI) site. Recently we received updates of sequencing progress from Joint Genome Institute (JGI) and we were able to locate, in this update, both the *aguA* and *xynA1* genes within the nucleotide sequence of a large contig, here designated as contig 1. The nucleotide positions of these genes in this contig are used in this figure. Nucleotide 956255 is the last coding nucleotide of *xynB* and nt 960933 is the first coding nucleotide of *aguA*. Nucleotide 2269148 is the first coding nucleotide of *xynA1* and nucleotide 2273536 is the last coding nucleotide for *xynA1*. Nucleotide 2377096 is the last nucleotide of this contig.

of genes encoding a multimodular GH10 xylanase containing SLH domains on the C-terminus and a secretion signal (XynA1) and homologs of the intracellular GH67 AguA and GH10 xylanase catalytic domain (XynA2) were used for the selection of candidates for further comparison. *G. stearothermophilus* T6 is included as a Firmicutes species that has been the most extensively characterized with respect to structure/function of GH10 xylanases and GH67 α-glucuronidase, as well as regulation of gene expression (Shulami et al., 1999, 2007). *Thermotoga maritima* MSB8 in a phylum distant to Firmicutes with respect to 16S rRNA sequence phylogeny contains a cell-associated multimodular xylanase (Liebl et al., 2008), as well as orthologs for the *aguA* and *xynA2* genes of the selected Firmicutes species. *Thermotoga maritima* MSB8 has also been evaluated for gene expression in response to substrate utilization (Chhabra et al., 2003). *Streptomyces avermitilis* is an example of a genome from the phylum Actinobacteria that has orthologs for all genes comprising the xylan regulon of *Paenibacillus* sp. JDR-2, although there are no sequences in the xylanases that are candidates for cell association. Table 1 presents a comparison of proteins and encoding genes for the secreted xylanase XynA1, intracellular AguA, intracellular XynA2, substrate-binding ABC transporter protein (Sbp), and the AraC response-regulator (AraC RR) protein for these different species.

The most highly conserved protein based upon amino acid sequence similarities is the GH67 AguA, followed by the xylanases, the substrate-binding protein, and then the transcriptional regulator, here represented by the AraC response regulator. The sizes of the proteins encoded by each gene type are also relatively constant, with the multimodular XynA1 represented only by its catalytic domain. The GC content of each gene is similar to the average GC content of the respective genome, indicating that these genes have evolved in their hosts over some time and providing evidence against recent horizontal transfer. Of the sequenced bacterial genomes in the genera evaluated here, all but one have a single copy of the *aguA* gene. *Geobacillus* Y412MC10 is the exception and has a paralog (GYMC10DRAFT_2223) with 61% identity to the one located within an aldouronate utilization gene cluster.

The secreted GH10 xylanase from *T. maritima* MSB8 lacks domains on the C-terminus that can be identified as SLH domains on the basis of sequence. The demonstration that *T. maritima* MSB8 secretes a GH10 xylanase associated with the outer membrane or toga through a specific sequence on the amino terminus (Liebl et al., 2008) provides evidence for another motif to serve a role similar to that proposed for the SLH domains of other xylanases. The XynA1 ortholog identified from the sequenced genome of *S. avermitilis* contains a CBM domain on the C-terminus, although it is not known whether this enzyme is associated with the cell surface.

Multimodular Cell-Associated GH10 Endoxylanases of Bacteria Containing *aguA*

Domain alignments of multimodular GH10 endoxylanases from bacteria that have genes encoding orthologs to a secreted GH10 XynA1-like xylanase with C-terminal SLH domains or other means of surface anchoring, an intracellular GH67 AguA, and an intracellular GH10 XynA2 are presented in Fig. 4.

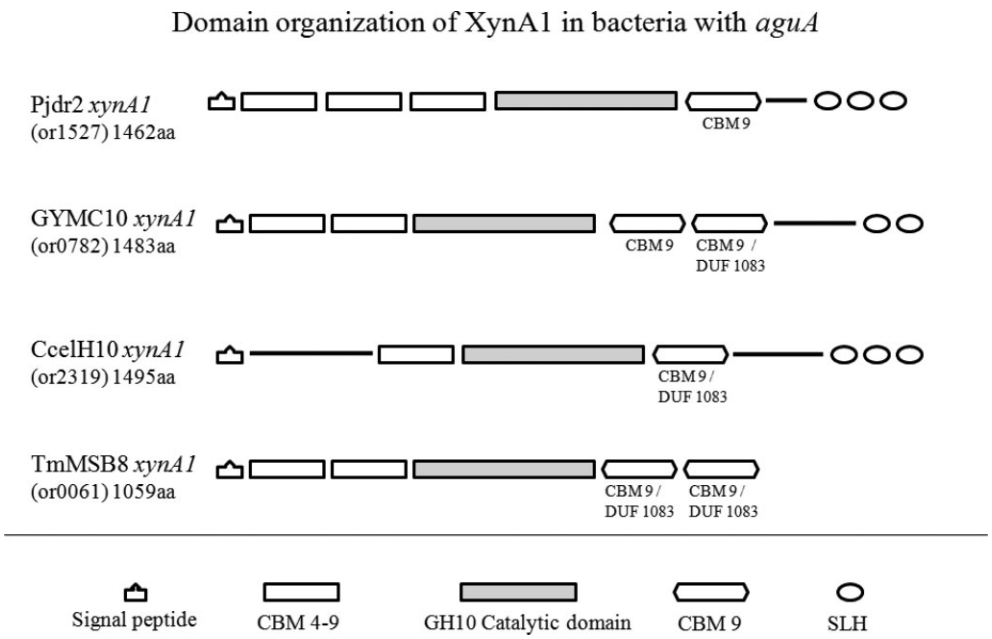

Figure 4. Domain organization of multimodular cell-associated GH10 endoxylanase (XynA1) from bacteria that have genes encoding a GH67 α-glucuronidase (AguA). Sequences of genes orthologous to XynA1 in *Paenibacillus* sp. JDR-2 were located by searching in IMG (Integrated Microbial genomes—http://img.jgi.doe.gov/cgi-bin/pub/main.cgi?page = home). Pjdr2: *Paenibacillus* sp. JDR-2; GYMC10: *Geobacillus* sp. Y412MC10; CcellH10: *Clostridium cellulolyticum* H10; TmMSB8: *Thermotoga maritima* MSB8.

In *Clostridium cellulolyticum* H10, there is a paralog to the *xynA1* Ccel_2319 depicted in Fig. 4. This paralog is the gene adjacent to Ccel_2319 (Ccel_2320), which, though smaller (encoding a protein of 1050 amino acids), encodes an additional CBM 4–9 domain. The extensive studies on the role of SLH domains in their association with bacterial cell surfaces via interaction with polysaccharides associated with the S-layer (Cava et al., 2004; Sara et al., 1998) support such a function for the same domains in many other glycohydrolases secreted by bacteria. Included would be those secreted by *Thermoanaerobacterium* spp. and other members of the Clostridiales (Ali et al., 2001; Brechtel & Bahl, 1998; Kosugi et al., 2002; Liu et al., 1996) as well species of *Caldocellulodisruptor* spp. (Ozedemir et al., 2012) and the thermophilic *Caldanaerobius polysaccharolyticus*, which also has a xylan-utilization gene cluster (Han et al., 2012). This list may be greatly expanded to include additional bacterial families and species with genes required for aldouronate utilization.

Genomic Organization for Orthologs of Genes Comprising the Xylan-Utilization Regulon of *Paenibacillus* sp. JDR-2

To gain insight into the functional arrangement of the genes comprising xylan-utilization regulons in different bacteria, completed or draft sequences were subjected to comparative analysis. The selection of genomes for comparison was based upon the same criteria used for comparison of the secreted multimodular GH10 endoxylanases in Fig. 4 and is presented in Fig. 5.

Distribution of *xynA1*, *xynA2*, and *aguA* Genes and Their Relationship to Transporter Genes

The distributions of the *xynA1* gene encoding the secreted multimodular xylanase are variable with respect to the *xynA2* and *aguA* genes, and have led to the distinction of two regions, one for *xynA1* and the second for *aguA* (Fig. 5a–c). The absence of secretion signals in *xynA2* and *aguA*, as well as any of the other genes within the aldouronate gene clusters, suggests that these all encode intracellular enzymes. The intracellular localization of XynA2 in *Paenibacillus* sp. JDR-2 is supported by the application of an activity assay that distinguishes XynA1 and XynA2 (Nong et al., 2009). For the Firmicutes species, the region in which the *xynA1* gene is found is always distal from the region that includes the *aguA* and *xynA2* as well as genes encoding ABC transporters and transcriptional regulators.

The relative locations of *aguA* and *xynA2* are also variable in the *aguA* regions, as are the relative positions of these genes to ABC transporters and transcriptional regulators. Most similar are *Paenibacillus* sp. JDR-2 (Fig. 5a) and *Geobacillus* Y412MC10 (Fig. 5b), for which the *aguA* gene is found adjacent to *xynA2*, both of which are immediately downstream from three genes encoding ABC transporter proteins: the substrate binding protein, permease 1, and permease 2 (*ugpB*, *lplB*, and *ugpE*, respectively). Notably absent from this group is a gene encoding an ATPase that is required for transporter function, indicating this activity may be supplied by a protein encoded by a distal gene, possibly constitutive if not coregulated with this cluster. Immediately upstream from the gene encoding the transporter proteins are two genes encoding transcriptional regulators, including a signal transduction protein of the histidine kinase family and a response regulator of the AraC/XylS family, respectively designated *yesM* and *yesN* in *Paenibacillus* sp. JDR-2 (Figs. 3 and 5a).

Thermotoga maritima (Fig. 5d) has an ortholog of *xynA1* (*xyl10A*) in a sequence proximal to *aguA* and an ortholog of *xynA2* (*xyl10B*) that is separated from *aguA* by several genes. Of the genomic profiles presented in Fig. 5, only *T. maritima* MSB8 (Fig. 5d) has a gene encoding an ATPase protein in the transporter cluster, and in fact has two such genes in both of the five subunit ABC transporters. The genes encoding the ATPase proteins are *or0057* and *or0058* in

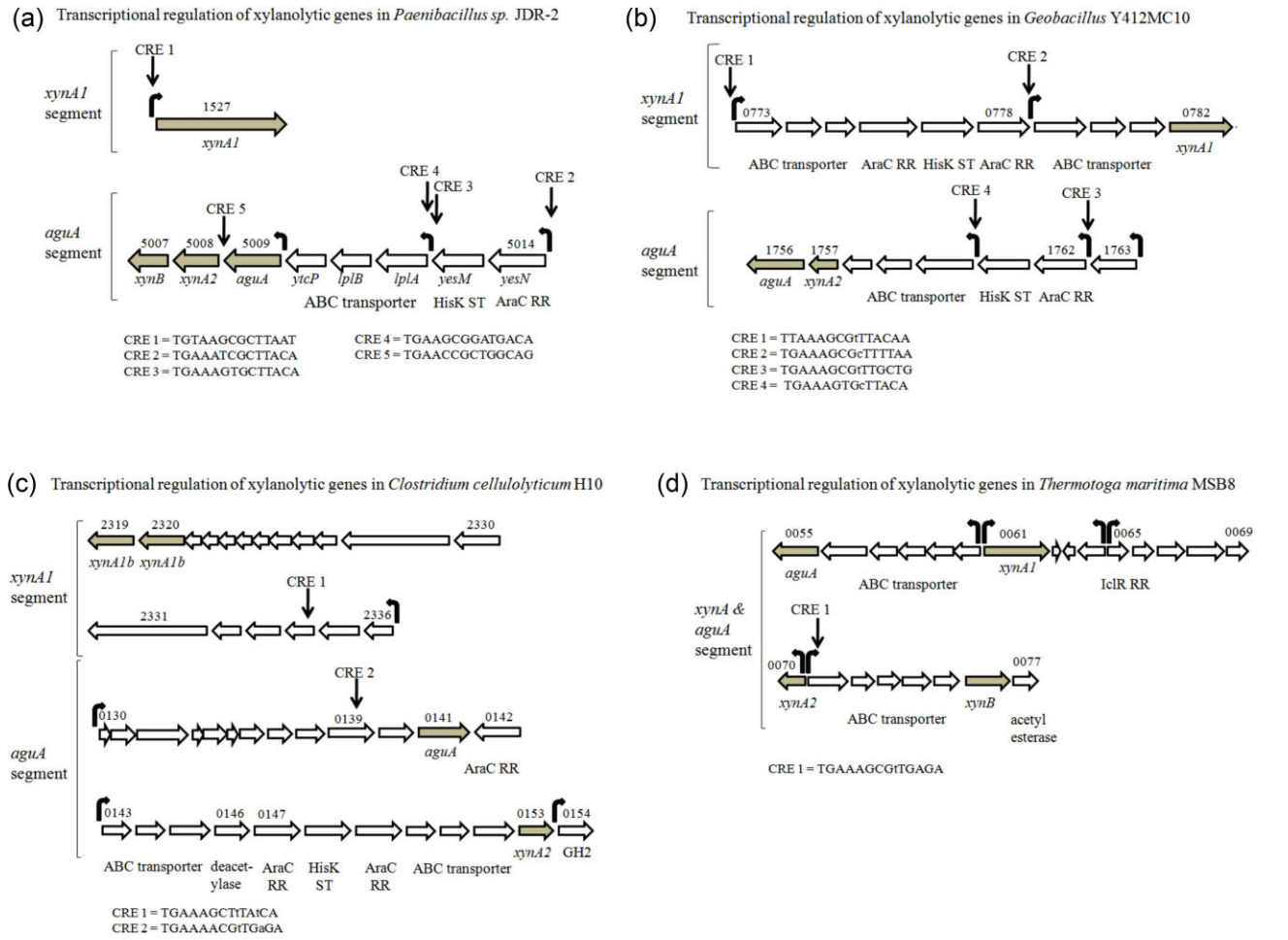

Figure 5. Linear representations of the genomic regions from each four bacteria containing genes encoding multimodular GH10 xylanases and GH67 α-glucuronidases. Nucleotide sequences of these organisms were obtained from IMG (Integrated Microbial genomes—http://img.jgi.doe.gov/cgi-bin/pub/main.cgi?page = home) and from NCBI (http://www.ncbi.nlm.nih.gov/). Promoters identified by the BPROM program (http://linux1.softberry.com/berry.phtml?topic = bprom&group = programs&subgroup = gfindb) with a score >4 and not located within the coding sequence are represented by arrows bent in the direction of transcription. Carbon repression element (CRE) sites were identified by the PPP program (Prokaryotic Pronoter Prediction—Groningen Biomolecular Sciences and Biotechnology Institute, Haren, the Netherlands). CRE sites, denoted by downward arrows, are identified in plus and minus strands in accordance with the direction of transcription of the genes in that region. Annotations of genes are as follows: *xynA1* encodes a multimodular GH10 xylanase; *xynB* encodes a β-xylosidase enzyme. (a) Xylanolytic gene organization and CRE sites in *Paenibacillus* sp. JDR-2. Data on nucleotide sequence and gene numbers on the *xynA1* segment and that on the *aguA* segment were obtained from the sequences NZ_ABKS01000003 and NZ_ABKS010000022, respectively, archived in National Center for Biotechnology Information (NCBI). CRE sites were determined as previously described (Chow et al., 2007). (b) Xylanolytic gene organization and CRE sites in *Geobacillus* Y412MC10. Data on nucleotide sequence and gene numbers on the *xynA1* segment and that on the *aguA* segment were obtained from the sequences NZ_ABRG01000001 and NZ_ABRG01000003, respectively, archived in NCBI. CRE 1 is at nucleotide 880514–18, CRE 2 is at nucleotide 888865–19, CRE 3 is at nucleotide 36948–62, and CRE 4 is at nucleotide 365873–86 of their respective archived sequences specified above. A second *aguA* is also found in this bacterium, but as the characteristic ABC transporter and transcriptional regulator were not found in its vicinity, it is not depicted in this diagram. (c) Xylanolytic gene organization and CRE sites in *Clostridium cellulolyticum* H10. Data on nucleotide sequence and gene numbers on the *xynA1* segment and that on the *aguA* segment were obtained from the sequence NC_011898, archived in NCBI. (d) Xylanolytic gene organization and CRE sites in *Thermotoga maritima* MSB8. *xynA1* and *aguA* are placed close to each other in this genome. Data on nucleotide sequence and gene numbers on these segments were obtained from NC_000853, archived in NCBI.

the ABC transporter gene cluster, just in front of *aguA*, and *or0074* and *or0075* in the ABC transporter between *xynA2* and *xynB* encoding a putative β-xylosidase.

Repression by the CcpA System. Identification of Carbon Repression Element Sites

The CcpA system has been defined for the repression of catabolic gene expression in genera and species within the Firmicutes as well as other bacterial phyla (Warner and Lolkema, 2003). CcpA, in complex with HPr phosphorylated at a specific serine residue, binds to a *cis*-acting catabolite carbon repression element (CRE) and blocks transcription of one or more genes within an operon (Zalieckas et al., 1998). The CRE sites have been defined as 14 base pairs that include the canonical sequence TG(T/A)AANCGNTN(A/T)CA in *G. stearothermophilus* and *Bacillus subtilis* (Cho and Choi, 1999). CRE sites have also been identified in genes within the xylan-utilization regulon of *Paenibacillus* sp. JDR-2 at locations in which they might be expected to account for the coordinate glucose repression of these genes (Chow et al., 2007). These sites are noted in the gene arrangements in Fig. 5a for *Paenibacillus* sp. JDR-2. CRE sites meeting

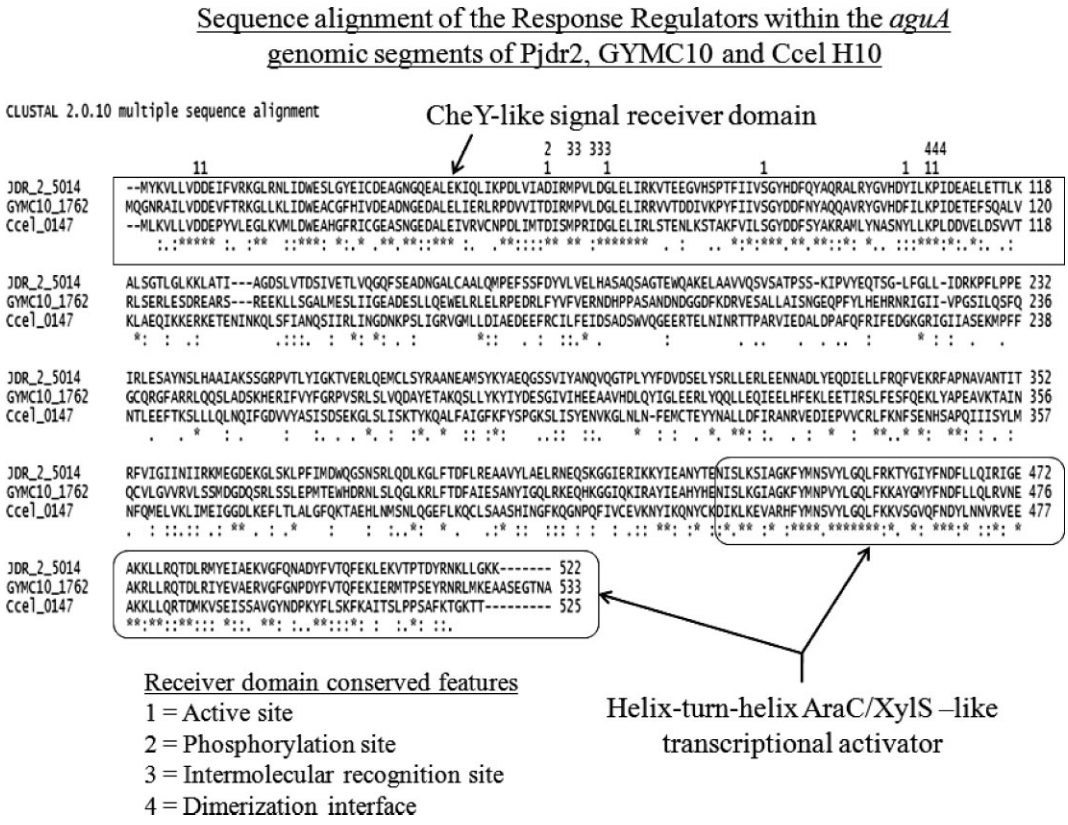

Figure 6. Identification of functional sequences in the AraC response regulators. The protein sequences aligned are gene 5014 of *Paenibacillus* sp. JDR-2 (JDR_2_5014), gene 1762 of *Geobacillus* Y416MC10 (GYMC10_1762), and gene 0142 of *Clostridium cellulolyticum* H10 (Ccel_0142). The number at the end of each line denotes the residue number of the amino acid at the end of that line. Asterisks below each column of three aligned residues represent identity, while colons and periods represent similarity. "1" identifies the residues that participate in the phosphorylation process. "2" indicates the aligned residues Asp55 (JDR_2_5014), Asp57 (GYMC10_1762), and Asp55 (Ccel_0142) that correspond to residue Asp57 of the AraC protein of *Escherichia coli* and are the sites of phosphorylation in this signal transduction system. Response regulators function as dimers. "3" and "4" indicate the residues that participate in the dimerization of the monomers.

the criteria of the canonical sequence are also identified in the gene arrangements provided for *Geobacillus* sp Y412MC10 (Fig. 5b), *C. cellulolyticum* H10 (Fig. 5c), and *T. maritima* MSB8 (Fig. 5d). Very similar arrangements were found for genes encoding response regulators, ABC transporters, AguA, and XynA2 for both *Paenibacillus* sp JDR-2 and *Geobacillus* sp Y412MC10, and the location of CRE sites are indicated within these regions. The gene arrangement in *C. cellulolyticum* is notably different, with CRE sites identified at locations within genes. All of the CRE sites in the xylan-utilization regulon in *Paenibacillus* sp. JDR-2 were located within or very near sites that contained strong promoters. These preceded two genes encoding transcriptional regulators, three genes encoding ABC transporters, three genes encoding intracellular glycohydrolases, all within the aldouronate-utilization cluster, and a site at the distal gene that encoded the secreted multimodular cell-associated GH10 endoxylanase, XynA1 (Chow et al., 2007). The arrangement of the genes in *Paenibacillus* sp. JDR-2 makes it attractive as a source of genes for the construction of synthetic operons. The CRE sites offer targets for mutations that render them incapable of binding activated CcpA complexes, and thereby insensitive to glucose repression. This may be useful in the development of biocatalysts able to simultaneously utilize cellulose and xylan and carry on SSF on a macroscale.

A CRE orthologous sequence was detected in the xylan-utilization regulon proposed for the genome of *T. maritima* MSB8. It is located within the substrate-binding protein of the ABC transporter, which preceded a *xynB* gene encoding a putative β-xylosidase. In *T. maritima* TM1200, a gene encoding proteins of low amino acid sequence similarity (32% identity) to the *ccpA* of *B. subtilis* subsp. *subtilis* 168 is found within the genome. However, HPr and HPK defined in the functions of carbon catabolite repression found in the Firmicutes were not found in *T. maritima*. It is not known whether the CcpA-mediated mechanism of glucose repression operates in *Thermotoga* spp. as it does in species of the Firmicutes. The putative CRE sites in *T. maritima* are less similar to the canonical sequences of *Bacillus* spp. for which they have been experimentally verified (Miwa et al., 2000).

Two-Component Signal Transduction Regulation

Two genes, *yesN* and *yesM*, are at the start of the alduronate-utilization gene cluster (Fig. 3). On the basis of Blastp, the *yesN* gene and *yesM* genes of *Paenibacillus* sp. JDR-2 encode a response regulator of the AraC/XylS family and a histidine kinase involved in signal transduction, respectively. Orthologs were found for all of the Firmicutes and also in *S. avermitilis*, but were not present in *T. maritima* MSB8. Amino acid sequence alignments of the yesN orthologs from *Paenibacillus* sp. JDR-2, *C. cellulolyticum* H10, and *Geobacillus* sp. Y412MC10 clearly identified the regulatory regions for CheY and AraC/XylS domains (Fig. 6). Although an ortholog

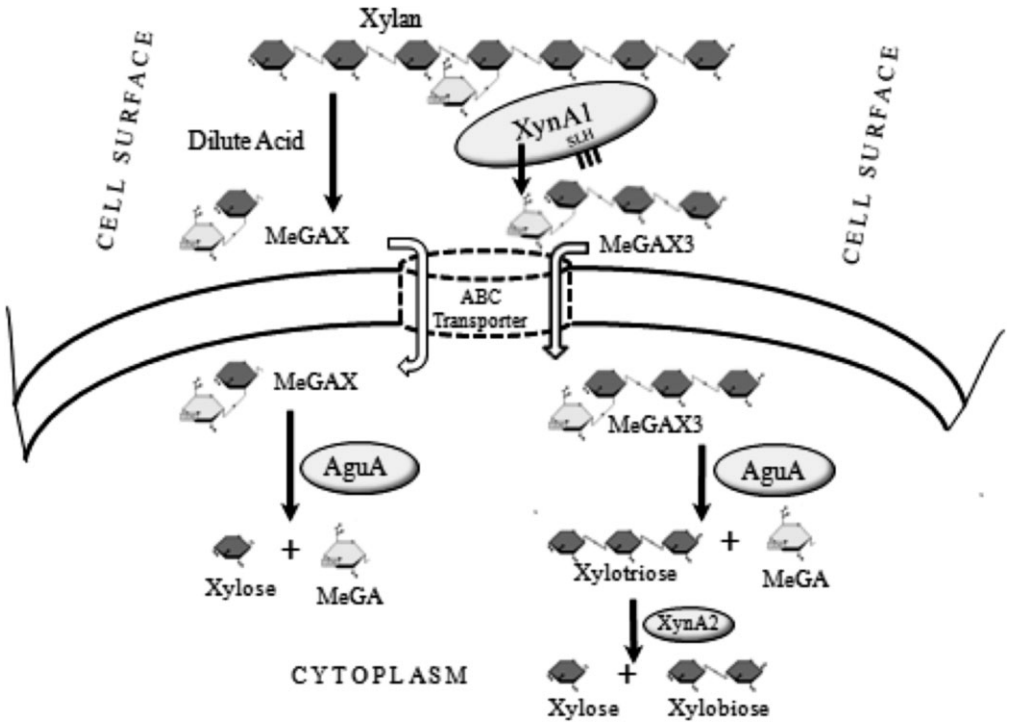

Figure 7. Diagram depicting cellular location of enzymes for coupling depolymerization, assimilation, and metabolism.

of the AraC response regulators found in Firmicutes could not be located in the xylan-utilization regulon in *T. maritima* MSB8, a gene encoding a transcriptional regulator of the Ic1R family was located at locus 0065 between the orthologous genes *xynA1* and *xynA2* (Fig. 5d).

The response regulators orthologous to the protein encoded by the *yesN* gene in the Firmicutes, represented by *Paenibacillus* sp. JDR-2, *Geobacillus* sp Y412MC10, and *C. cellulolyticum* H10, are ~520 amino acids and include a CheY-like receiver domain in the C-terminus and an HTH AraC/XylS domain in the N-terminus. In this respect, the protein encoded by the *yesN* gene is functionally similar to the AraC, XylS, and RhaR genes that have been well studied in *E. coli*, although like orthologs in other Gram-negative bacteria, these are smaller and are comprised of ~300 amino acids (Gallegos et al., 1997). The *xynR* gene in *Prevotella bryantii* B14 encodes an 800 amino acid multidomain protein that includes the response regulator and signal transduction functions ascribed to the proteins encoded by the yesN and yesM in *Paenibacillus* sp. JDR-2, and molecular studies have demonstrated that this regulates the expression of genes comprising a xylanase operon (Miyazaki et al., 2003). The co-induction of all of the genes ascribed to the xylan-utilization regulon in *Paenibacillus* sp. JDR-2 during growth on either MeGAX$_n$ or xylose and corepression of these same genes during growth on glucose attest to the significance of the response regulators and the associated signal transduction proteins in this regulon. *Geobacillus* sp. Y412MC10 and *C. cellulolyticum* H10, as well as *Paenibacillus* sp. JDR-2, have a number of paralogs in their respective genomes that are proximal to glycohydrolases. It is likely that the *yesN* and *yesM* orthologs would be required for the regulated expression of the other genes encoding the ABC transporter proteins and glycohydrolases and would be required for the construction of functional synthetic operons.

Potential for Developing a Single Biocatalyst to Efficiently Convert Xylose From Glucuronoxylans to Fuels or Chemicals

Paenibacillus sp. JDR-2 cultures demonstrate a greater rate of growth and produce more biomass with MeGAX$_n$ as substrate than with comparable mass quantities of monosaccharides, for example xylose, glucose, or glucuronate (St John et al., 2006a). The ratios of the maximum rates of utilization for MeGAX$_n$ to MeGAX$_3$ (the aldouronate generated by the action of the XynA1 on MeGAX$_n$) or MeGX (the aldouronate generated by the dilute acid hydrolysis of MeGX$_n$) were 2.75 and 2.5, respectively. The rates of utilization of MeGAX$_n$, MeGAX$_3$, and MeGX were proportional to the rates of biomass formation and to the amount of biomass formed (Nong et al., 2009). The absence of detectable oligosaccharides in the medium during growth supports a process in which the assimilation of products occurs rapidly as the products are formed. The anchoring of the multimodular GH10 endoxylanase to the cell surface through three SLH domains provides a rational basis for generation of the products at sites where, with minimal diffusion, they are recognized by the substrate-binding proteins and efficiently delivered into the cell. Assuming the delivery process relies on group transfers of oligosaccharides, the assimilation consumes fewer ATP equivalents per xylose than would be required for mono- or disaccharides. Thus, the process provides bioenergetic coupling as a consequence of metabolism and is kinetically favored by the proximal location of depolymerization to assimilation. The diagram in Fig. 7 presents the overall process of the utilization of MeGAX$_3$ enzymatically derived from polymeric MeGAX$_n$ as well as MeGX, a product of dilute acid hydrolysis.

The formation of xylotriose and xylobiose catalyzed by the cell-associated XynA1 may provide a similar advantage to that provided by the generation of MeGAX$_3$ as none of these products are

detected in the medium during the rapid phase of growth (Nong et al., 2009; St John et al., 2006a). It is not known if the transporters that assimilate the aldouronates also transport the xylobiose and xylotriose. The role of the AguA to cleave the α-1,2-glycosidic bond linking MeGA to xylose as in the case of MeGX, as well when linking to the nonreducing terminus of xylotriose in $MeGAX_3$, would suggest that the substrate-binding protein encoded by the *lplA* gene in Fig. 3 would discriminate between an aldouronate and a neutral oligoxyloside (Nong et al., 2009).

A significant number of bacteria contain genes encoding multimodular GH10 endoxylanases with SLH domains (St John et al., 2006a), and this number is increasing as more genomes are sequenced. The common substitution of α-linked MeGA at the 2-*O* position on xyloses within the backbone of the $MeGAX_n$ presents a metabolic challenge to microorganisms that rely on hemicellulose as a principal substrate. It is thus somewhat surprising that so few of these also contain *aguA* genes encoding GH67 AguAs. The removal of MeGA from the $MeGAX_n$ by secreted AguAs has been demonstrated with fungal enzymes (Tenkanen & Siika-aho, 2000). Most recently an extracellular AguA has been purified from cultures of the yeast *Pichia stipitis* with the ability to catalyze the extracellular removal of the MeGA from $MeGAX_n$ (Ryabova et al., 2009). The gene encoding this enzyme has been located in the sequenced genome of *P. stipitis*, and its amino acid sequence and catalytic properties have led to its classification in glycohydrolase family 115. The expression of a gene encoding a GH115 AguA was observed in the transcriptome study of *Paenibacillus* sp. JDR-2 when grown on $MeGAX_n$ (Sawhney et al., 2016). Unlike the *Paenibacillus* sp. JDR-2 GH67 AguA, which was specific for the $MeGAX_3$ in which MeGA is linked to the nonreducing terminal xylose in the aldouronate generated by the action of the GH10 xylanase, the recombinant GH115 from *Paenibacillus* sp. JDR-2 released MeGA from $MeGAX_3$ in which the linked xylose was other than the reducing terminus (Rhee et al., 2017). This enzyme also released MeGA from polymeric $MeGAX_n$ as well as the aldouronates generated by the action of GH11 and GH30 enzymes on $MeGAX_n$. The lack of SLH domain sequences in the genes encoding secreted xylanases other than the multimodular secreted XynA1 suggests that they may have a less significant role in the extracellular processing of $MeGAX_n$.

The cell-associated multimodular GH10 xylanases along with intracellular GH67 AguAs and the regulatory systems for promoting assimilation and metabolism of the products of depolymerization provide for the development of biocatalysts for converting hemicellulosic biomass to biofuels and chemicals. *Paenibacillus* sp. JDR-2 also has systems in which cell-associated multimodular β-1,3–1,4-glucanases (Chow et al., 2016) and amylases (Pan, 2014) participate in the efficient processing and metabolism of soluble glucans, adding to the components of plant biomass that can be efficiently processed (Sawhney et al., 2016).

Many cellulolytic and xylanolytic bacterial species in the Firmicutes have evolved systems that rapidly assimilate the products of depolymerization, providing a more energetically efficient and rapid rate of bioconversion, mitigating the need for harsh chemical pretreatments to release fermentable substrates. The system in *Paenibacillus* sp. JDR-2 and the orthologous systems described here offer a potential for the development of more efficient biocatalysts to process underutilized sources of plant biomass. Development of new biocatalysts through downstream engineering for fermentation to targeted products offers a promising path for industrial production of fuels and chemicals. On the other hand, the xylan-utilization regulons described here may be used to engineer genetically malleable species for production of targeted biobased products from these resources.

Acknowledgments

The studies presented here resulted from an extensive, long-term collaboration with the research laboratories of Professors Lonnie O. Ingram and Keelnathan T. Shanmugam in the Department of Microbiology and Cell Science, at the University of Florida which focused on the development of bacterial biocatalysts for converting lignocellulosic biomass to fuels and chemicals. F.J.S.J. would like to thank the USDA Forest Service, Forest Products Laboratory, Institute for Microbial and Biochemical Technology for ongoing financial support. We appreciate the assistance provided by Christa Pennacchio for managing the RNA sequencing project and also the other members of the Joint Genome Institute, Walnut Creek, CA, for conducting the DNA and RNA sequencing work, which was supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

Funding

This research was supported by the US Department of Energy grants DE FC36-99GO10476 and DE FC36-00GO10594; the Consortium for Plant Biotechnology Research Project GO12026-198 (DE FG36-02GO12026); the Institute of Food and Agricultural Sciences, University of Florida Experiment Station, CRIS Project MCS 3763; Biomass Research & Development Initiative competitive grant no. 2011-10006-30358 from the USDA National Institute of Food and Agriculture; and the Florida Energy Systems Consortium, State University System of FL, project no. 00077818.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability

All data presented in this review is in the public domain and can be retrieved through the National Center for Biotechnology Information (NCBI) web search tools (https://www.ncbi.nlm.nih.gov/search/).

References

Agbogbo, F. K. & Caward-Kelly, G. (2008). Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast. *Biotechnology Letters*, 30, 1515–1524. https://doi.org/10.1007/s10529-008-9728-z.

Ali, M. K., Kimura, T., Sakka, K., & Ohmiya, K. (2001). The multidomain xylanase XynlOB as a cellulose-binding protein in *Clostridium stercorarium*. *FEMS Microbiology Letters*, 198, 79–83. https://doi.org/10.1111/j.1574-6968.2001.tb10622.x.

Bergquist, P. L., Morgan, H. W., & Saul, D. (2014). Selected enzymes from extreme thermophiles with applications to biotechnology. *Current Biotechnology*, 3, 45–59. https://doi.org/10.2174/2211550102999131230150918.

Biely, P., Vršanská, M., Tenkanen, M., & Kluepfel, D. (1997). Endo-beta-1,4-xylanase families: differences in catalytic properties. *Journal of Biotechnology*, 57, 151–166. https://doi.org/10.1016/s0168-1656(97)00096-5.

Biely, P, Malovíková, A, Hirsch, J, Morkeberg Krogh, K. B., & Ebringerová, A. (2015). The role of the glucuronoxylan carboxyl groups in the action of endoxylanases of three glycoside hydrolase families: A study with two substrate mutants. *Biochimica et Biophysica Acta - General Subjects*, 1850, 2246–2255. https://doi.org/10.1016/j.bbagen.2015.07.003.

Boranston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: Fine-tuning polysaccha-

ride recognition. *Biochemical Journal*, 382, 769–781. https://doi.org/10.1042/BJ20040892.

Brechtel, E. & Bahl, H. (1999). In *Thermoanaerobacterium thermosulfurigenes* EM1 S-layer homology domains do not attach to peptidoglycan. *Journal of Bacteriology*, 181, 5017–5023. https://doi.org/10.1128/JB.181.16.5017-5023.1999.

Cava, F., de Pedro, M. A., Schwarz, H., Henne, A., & Berenguer, J. (2004). Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria. *Molecular Microbiology* 52, 677–690. https://doi.org/10.1111/j.1365-2958.2004.04011.x.

Chhabra, S. R., Shockley, K. R., Conners, S. B., Scott, K. L., Wolfingers, R. D., & Kelly, R. M. (2003). Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *Journal of Biological Chemistry*, 278, 7540–7552. https://doi.org/10.1074/jbc.M211748200.

Cho, S.-G. & Choi, Y.-J. (1999). Catabolite repression of the xylanase gene (*xynA*) expression in *Bacillus stearothermophilus* No. 236 and *B. subtilis*. *Bioscience, Biotechnology, and Biochemistry*, 63, 2053–2058. https://doi.org/10.1271/bbb.63.2053.

Chow, V., Nong, G., & Preston, J. F. (2007). Structure, function and regulation of the aldouronate-utilization gene cluster from *Paenibacillus* sp. JDR-2. *Journal of Bacteriology*, 189, 8863–8870. https://doi.org/10.1128/JB.01141-07.

Chow, V., Nong, G., St John, F. J., Rice, J. D., Dickstein, E., Chertkov, O., Bruce, D., Detter, C., Brettin, T., Han, J., Woyke, T., Pitluck, S., Nolan, M., Pati, A., Martin, J., Copeland, A., Land, M.L., Goodwin, L., Jones, J.B., … Preston, J. F. (2012). Complete genome sequence of *Paenibacillus* sp. strain JDR. *Standards in Genomic Sciences*, 6, 1. https://doi.org/10.4056/sigs.2374349.

Chow, V., Kim, Y.S., Rhee, M.S., Sawhney, N., St John, F. J., Nong, G., & Preston, J. F., (2016). Definition of a 1,3-1,4- β-glucan-utilization regulon in *Paenibacillus* sp. JDR-2. *Applied and Environmental Microbiology*, 82, 1789–1798. https://doi.org/10.1128/AEM.03526-15.

Chundawet, S. P. S., Beckham, G. T., Himmel, M. E., & Dale, M. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annual Review of Chemical and Biomolecular Engineering*, 2, 121–145.

Collins, T., Gerday, C., & Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews*, 29, 3–23. https://doi.org/10.1016/j.femsre.2004.06.005.

Dien, B. S., Cotta, M. A., & Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: current status. *Applied Microbiology and Biotechnology*, 63, 258–266. https://doi.org/10.1007/s00253-003-1444-y.

Dien, B. S., Jung, H. J. G., Vogel, K. P., Casler, M. D., Lamb, J. F. S., Iten, L., Mitchell, R. B., & Sarath, G. (2006). Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass and Bioenergy*, 30, 880–891. https://doi.org/10.1016/j.biombioe.2006.02.004.

Gallegos, M.-T., Schleif, R., Bairoch, A., Hofmann, K., & Ramos, J. L. (1997). AraC/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews*, 61, 393–410. https://doi.org/10.1128/mmbr.61.4.393-410.1997.

Han, Y., Argawal, V., Dodd, D., Kim, J., Bae, B., & Mackie, R. I. (2012). Xylan utilization by thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *Journal of Biological Chemistry*, 287, 34946–34960. https://doi.org/10.1074/jbc.M112.391532.

Himmel, M., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., & Foust, T. D. (2007). Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, 315, 804–807. https://doi.org/10.1126/science.1137016.

Hurlbert, J. C. & Preston, J. F. (2001). Functional characterization of a novel xylanase from corn strains of *Erwinia chrysanthemi*. *Journal of Bacteriology*, 183, 2093–2100. https://doi.org/10.1128/JB.183.6.2093-2100.2001.

Ingram, L. O., Conway, T., Clark, D. P., Sewell, G. W., & Preston, J. F. (1987). Genetic engineering of ethanol production in *Escherichia coli*. *Applied and Environmental Microbiology*, 53, 2420–2425. https://doi.org/10.1128/aem.53.10.2420-2425.1987.

Ingram, L. O., Aldrich, H. C., Borges, A. C. C., Causey, T. B., Martinez, A., Morales, F., Saleh, A., Underwood, S. A., Yomano, L. P., York, S. W., Zaldivar, J., & Zhou, S. (1999). Enteric bacterial catalysts for fuel ethanol production. *Biotechnology Progress*, 15, 855–866. https://doi.org/10.1021/bp9901062.

Ito, Y., Tomita, T., Roy, N., Nakano, A., Sugawara-Tomita, N., Watanabe, S., Okai, N., Abe, N., & Kamio, Y. (2003). Cloning, expression, and cell surface localization of *Paenibacillus* sp. strain W-61 xylanase 5, a multidomain xylanase. *Applied and Environmental Microbiology*, 69, 6969–6978. https://doi.org/10.1128/AEM.69.12.6969-6978.2003.

Jacobs, A., Larsson, P. T., & Dahlman, O. (2001). Distribution of uronic acids in xylans from various species of soft- and hardwood as determined by MALDI mass spectrometry. *Biomacromolecules*, 2, 979–990. https://doi.org/10.1021/bm010062x.

Jarboe, L. R., Grabar, T. B., Yomano, L. P., Shanmugan, K. T., & Ingram, L. O. 2007. Development of ethanologenic bacteria. *Advances in Biochemical Engineering and Biotechnology*, 108**,** 237–261. https://doi.org/10.1007/10_2007_068.

Jeffries, T. W. & Jin, Y. S. (2004). Metabolic engineering for improved fermentation of pentoses by yeasts. *Applied Microbiology and Biotechnology*, 63, 495–509. https://doi.org/10.1007/s00253-003-1450-0.

Jeffries, T. W. (2006). Engineering yeasts for xylose metabolism. *Current Opinion in Biotechnology*, 17, 320–326. https://doi.org/10.1016/j.copbio.2006.05.008.

Kosugi, A., Murashima, K., Tamaru, Y., & Doi, R. H. (2002). Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. *Journal of Bacteriology*, 184, 884–888. https://doi.org/10.1128/jb.184.4.884-888.2002.

Kuhad, R. C., Singh, A., & Eriksson, K. E. (1997). Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Advances in Biochemical Engineering and Biotechnology* 57, 45–125. https://doi.org/10.1007/BFb0102072.

Liebl, W., Winterhalter, C., Baumeister, W., Armbrecht, M., & Valdez, M. (2008). Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. *Journal of Bacteriology*, 190, 1350–1358. https://doi.org/10.1128/JB.01149-07.

Liu, C. & Wyman, C. E. (2005). Partial flow of compressed hot water through corn stover to enhance hemicellulose sugar recoveries and enzymatic digestibility of cellulose. *Bioresource Technology*, 96, 1978–1985. https://doi.org/10.1016/j.biortech.2005.01.012.

Liu, S.-Y., Gherardini, F. C., Matuscheck, M., Bahl, H., & Wiegel, J. (1996). Cloning, sequencing, and expression of a gene encoding a large S-layer –associated endoxylanase from *Thermoanaerobacterium* sp. strain JW/SL. *Journal of Bacteriology*, 178, 1539–1547. https://doi.org/10.1128/jb.178.6.1539-1547.1996.

Lloyd, T. A. & Wyman, C. E. (2005). Combined sugar yield for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource Technology*, 96, 1967–1977. https://doi.org/10.1016/j.biortech.2005.01.011.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 66, 506–577. https://doi.org/10.1128/MMBR.66.3.506-577.2002.

Mai, V. & Wiegel, J. (2000). Advances in the development of a genetic system for *Thermoanerobacterium* spp.: Expression of genes encoding hydrolytic enzymes, development of a second shuttle vector and integration of genes into the chromosome. *Applied and Environmental Microbiology*, 66, 4817–4821. https://doi.org/10.1128/AEM.66.11.4817-4821.2000.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., & Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *The EMBO Journal*, 19, 4473–4484. https://doi.org/10.1093/emboj/19.17.4473.

Miwa, Y., Nakata, A., Ogiwara, A., Yamamotto, M., & Fujita, Y. (2000). Evaluation and chacterization of catabolite-responsive elements (*cre*) of *Bacillus subtilis*. *Nucleic Acids Research*, 28, 1206–1210.

Miyazaki, K, Miyamoto, H., Mercer, D. K., Hirase, T., Martin, J. C., Kojima, Y., & Flint, H. J. (2003). Involvement of the multidomain regulatory protein XynR in positive control of xylanase gene expression in the ruminal anaerobe *Prevotella bryantii* B_14. *Journal of Bacteriology*, 158, 2219–2226. https://doi.org/10.1128/JB.185.7.2219-2226.2003.

Nong, G., Rice, J. D., Chow, V., & Preston, J. F. (2009). Aldouronate utilization in *Paenibacillus* sp. JDR-2: Physiological and enzymatic evidence for coupling of extracellular depolymerization with intracellular metabolism. *Applied and Environmental Microbiology*, 75, 4410–4418. https://doi.org/10.1128/AEM.02354-08.

Notenboom, V., Boraston, A. B., Kilburn, D. G., & Rose, D. R. (2001). Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry*, 40, 6248–6256. https://doi.org/10.1021/bi0101704.

Ozedemir, I, Blumer-Schutte, S. E., & Kelly, R. M. (2012). S-layer domain protein Csas 0678 and Csac 2722 are implicated in plant polysaccharide deconstruction by extremely thermophilic *Caldocellulodisruptor saccharolyticum*. *Applied and Environmental Microbiology*, 78, 768–777. https://doi.org/10.1128/AEM.07031-11.

Pan, L. (2014). *Starch utilization in Paenibacillus Sp. strain JDR-2: Depolymerization catalyzed by cell-associated alpha-amylase*. *Master's thesis*. University of Florida.

Preston, J. F., Hurlbert, J. C., Rice, J. D., Ragunathan, A., & St John, F. J. (2003). Microbial strategies for the depolymerization of glucuronoxylan: Leads to the biotechnological applications of endoxylanases. Application of enzymes to lignocellulosics. In S. D. Mansfield and J. N. Saddler (Eds.), *American Chemical Symposium Series No. 855* (pp. 191–210), ACS Publications.

Rhee, M. S., Sawhney, N., Kim, Y. S., Rhee, H. J., Hurlbert, J. C., St John, F. J., Nong, G., Rice, J. D., & Preston, J. F. (2017). GH115 α-glucuronidase and GH11 xylanase from *Paenibacillus* sp. JDR-2: Potential roles in processing glucuronoxylans. *Applied Microbiology and Biotechnology*, 101, 1465–1476. https://doi.org/10.1007/s00253-016-7899-4.

Ryabova, O., Vrsanaka, M., Kaneko, S., van Zyl, W. H., & Biely, P. (2009). A novel family of hemicellulolytic α-glucuronidase. *FEBS Letters*, 583, 1457–1462. https://doi.org/10.1016/j.febslet.2009.03.057.

Saha, B.C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*, 30, 279–291. https://doi.org/10.1007/s10295-003-0049-x.

Sára, M., Egelseer, E. M., Dekitsch, C., & Sleytr, U. B. (1998). Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-Layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. *Journal of Bacteriology*, 180, 6780–6783. https://doi.org/10.1128/JB.180.24.6780-6783.1998.

Sawhney, N., Crooks, C., St John, F. J., & Preston, J. F. (2015). Transcriptomic analysis of xylan utilization systems in *Paenibacillus* sp. JDR-2. *Applied and Environmental Microbiology*, 81, 1490–1501. https://doi.org/10.1128/AEM.03523-14.

Sawhney, N., Crooks, C., Chow, V., Preston, J. F., & St John, F. J. (2016). Genomic and transcriptomic analysis of carbohydrate utilization by *Paenibacillus* sp. JDR-2: Systems for processing plant polysaccharides. *Bmc Genomics*, 17, 131. https://doi.org/10.1186/s12864-016-2436-5.

Sedlak, M. & Ho, N. W. (2004). Production of ethanol from cellulosic biomass hydrolysates using genetically engineered *Saccharomyces* yeast capable of cofermenting glucose and xylose. *Applied Biochemistry and Biotechnology*, 114, 403–416. https://doi.org/10.1385/abab:114:1-3:403.

Shanmugam, K. T. & Ingram, L. O. (2008). Engineering biocatalysts for production of commodity chemicals. *Journal of Molecular Microbiology and Biotechnology*, 15, 8–15. https://doi.org/10.1159/000111988.

Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., Hogsett, D. A., & Lynd, L. R. (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol in high yield. *Proceedings of the National Academy of Sciences*, 105, 13769–13774. https://doi.org/10.1073/pnas.0801266105.

Shulami, S., Gat, O., Sonenshein, A. L., & Shoham, Y. (1999). The glucuronic acid-utilization gene cluster from *Bacillus stearothermophilus* T-6. *Journal of Bacteriology*, 181, 3695–3704. https://doi.org/10.1128/JB.181.12.3695-3704.1999.

Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L., & Shoham, Y. (2007). A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Applied and Environmental Microbiology*, 73, 874–884. https://doi.org/10.1128/AEM.02367-06.

St John, F. J., Rice, J. D., & Preston, J. F. (2006). *Paenibacillus* sp. JDR-2 and $XynA_1$: A novel system for methylglucuronoxylan utilization. *Applied and Environmental Microbiology*, 72, 1496–1506. https://doi.org/10.1128/AEM.72.2.1496-1506.2006.

St John, F. J., Rice, J. D., & Preston, J. F. (2006). Characterization of XynC from *Bacillus subtilis* subspecies *subtilis* strain 168 and analysis of its role in depolymerization of glucuronoxylan. *Journal of Bacteriology*, 188, 8617–8626. https://doi.org/10.1128/JB.01283-06.

St John, F. J., González, J. M., & Pozharski, E. (2010). Consolidation of glycosyl hydrolase family 30: A dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Letters*, 584, 4435–4441. https://doi.org/10.1016/j.febslet.2010.09.051.

St John, F. J., Preston, J. F., & Pozharski, E. (2012). Novel structural features of xylanase A1 from *Paenibacillus* sp. JDR-2. *Journal of Structural Biology*. 180, 303–311. https://doi.org/10.1016/j.jsb.2012.09.007.

Straub, C. T., Khatibi, P. A., Wang, J. P., Conway, J. M., Williams-Rhaesa, A. M., Peszlen, I. O., Chiang, V. L., Adams, M. W. W., & Kelly, R. M. (2019). Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. *Nature Communications*, 10, 3548. https://doi.org/10.1038/s41467-019-11376-6.

Tenkanen, M. & Siika-aho, M., (2000). An α-glucuronidase of *Shizophyllum commune* acting on polymeric xylan. *Journal of Biotechnology*, 78, 149–161. https://doi.org/10.1016/s0168-1656(99)00240-0.

van Maris, A. J. A., Abbott, D. A., Bellissimi, E., van den Brink, J., Kuyper, M., Luttik, M. A. H., Wisselink, H. W., Scheffers, W. A., van Dijken, J. P., & Pronk, J. T. (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek*, 90, 391–418. https://doi.org/10.1007/s10482-006-9085-7.

Vršanská, M., Kolenová, K., Puchart, V., & Biely, P. (2007). Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*. *FEBS Journal*, 274, 1666–1677. https://doi.org/10.1111/j.1742-4658.2007.05710.x.

Warner, J. B. & Lolkema, J. S. (2003). CcpA-dependent carbon catabolite repression in bacteria. *Microbiology and Molecular Biology Reviews*, 67, 475–490. https://doi.org/10.1128/MMBR.67.4.475-490.2003.

Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., & Lee, Y. Y. (2005). Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, 96, 1959–1966. https://doi.org/10.1016/j.biortech.2005.01.010.

Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M., & Bolam, D. N. (2001). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: The role of conserved amino acids in ligand binding. *Biochemistry*, 40, 9167–9176. https://doi.org/10.1021/bi0106742.

Zalieckas, J. M., Wray, L. V., & Fisher, S. H. (1998). Expression of the *Bacillus subtilis acsA* gene: Position and sequence context affect *cre*-mediated carbon catabolite repression. *Journal of Bacteriology*, 180, 6649–6654. https://doi.org/10.1128/JB.180.24.6649-6654.1998.

Zhang, Y.-H. P. & Lynd, L. R. (2005). Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation. *Proceedings of the National Academy of Sciences*, 102, 7321–7325. https://doi.org/10.1073/pnas.0408734102.

Zhao, G., Ali, E., Sakka, M., Kimura, T., & Sakka, K. (2006). Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. *Applied Microbiology and Biotechnology*, 70, 464–469. https://doi.org/10.1007/s00253-005-0079-6.

Zhao, G., Li, H., Wamalwa, B., Sakka, M., Kimura, T., & Sakka, K. (2006). Different binding specificities of S-layer homology modules from *Clostridium thermocellum* AncA, Slp1, and Slp2. *Bioscience, Biotechnology, and Biochemistry*, 70, 1636–1641. https://doi.org/10.1271/bbb.50699.